

Football Match Result Prediction

Machine Learning Project | Bar-Ilan University

Executive Summary

This project applies machine learning techniques to predict football match outcomes (Home Win, Draw, Away Win) based on extensive match statistics from the Kaggle Football Database. Using preprocessing, feature engineering, statistical testing, and model tuning, we developed a production-ready XGBoost classifier achieving high accuracy and stability.

1. Introduction

The dataset, available at: <https://www.kaggle.com/datasets/technika148/football-database>, contains football match records from Europe's top leagues. Each record includes over 100 features such as goals, assists, cards, xG, and more. The goal is to predict the match result: Home win (H), Draw (D), or Away win (A).

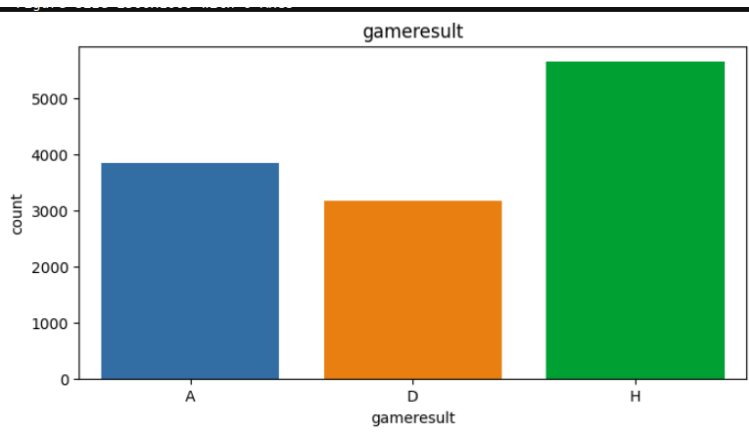
2. Objective

Develop a machine learning model to classify the match outcome using match statistics.

3. Dataset Overview

The dataset includes 12000+ matches and over 100 features.

- Target Variable: gameresult
- Feature Types: numeric (e.g., xG, shots), categorical (e.g., teamID), and binary (e.g., red cards).



NOTE: This is a placeholder preview of the full protocol PDF.

The complete structured PDF with all plots and statistical tests will be generated next.